Nanostructure and molecular mechanics of spider dragline silk protein assemblies

Sinan Keten[1], Markus J. Buehler[1,2,3†]

[1] Laboratory for Atomistic and Molecular Mechanics, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave. Room 1-235A&B, Cambridge, MA, USA

[2] Center for Materials Science and Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA, USA

[3] Center for Computational Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, Massachusetts 02139, USA

[†] Corresponding author: E-mail: mbuehler@MIT.EDU, phone: +1-617-452-2750, fax: +1-617-324-4014, lab URL: http://web.mit.edu/mbuehler/www/

Summary: Spider silk is a self-assembling biopolymer that outperforms most known materials in terms of its mechanical performance. While experimental studies have shown that the molecular structure of silk proteins has a direct influence on the stiffness, toughness and failure strength of silk, no molecular level analysis of the nanostructure and associated mechanical properties of silk assemblies have been reported. Here we report atomic-level structures of MaSp1 and MaSp2 proteins from the *N. Clavipes* spider dragline silk sequence, obtained using replica exchange molecular dynamics, and subject these structures to mechanical loading for a detailed nanomechanical analysis. The structural analysis reveals that poly-alanine regions in silk predominantly form distinct and orderly beta-sheet crystal domains, while disorderly regions are formed by glycine rich repeats that consist of 3_1-helix type structures and beta-turns. Our structural predictions are validated against experimental data based on dihedral angle pair calculations presented in Ramachandran plots, alpha-carbon atomic distances, as well as secondary structure content. A selection of structures is subject to mechanical loading, which reveals that distinctly different hydrogen bonded regions and the type of secondary structure control the mechanical response of silk at the nanoscale. Both structural and mechanical characterization results show excellent agreement with available experimental evidence. Our findings set the stage for extensive molecular-level investigations of silk, which may contribute towards an improved understanding of the source of the strength and toughness of this biological superfiber.

S. Keten, M.J. Buehler, "Nanostructure and molecular mechanics of dragline spider silk protein assemblies", Journal of the Royal Society Interface, Vol. 7(53), pp. 1709-1721, 2010

Keywords: Biological material, nanostructure, molecular modeling, materiomics, spider silk, deformation, failure

1. INTRODUCTION

Chemistry and nanoscale features of biological materials are crucial to understand the source of their mechanical properties such as strength, failure mechanisms and elasticity (Fratzl & Weinkamer 2007). In protein materials, the primary structure of macromolecules, consisting of a linear sequence of individual amino acids describes the chemical specificity of the interactions at the molecular level, which give rise to formation and failure characteristics of the material (Buehler & Yung 2009; LeDuc & Robinson 2007). Spider silk is an extraordinary material that surpasses most synthetic fibers in terms

of toughness through a balance of ultimate strength and extensibility (Becker et al. 2003; Shao & Vollrath 2002; Simmons et al. 1996; Termonia 1994; Vollrath & Knight 2001). The source of silk's unique properties has been attributed to the specific secondary structures of proteins found in the repeating units of spider silk proteins (Hayashi et al. 1999), which assemble into a hierarchical structure as shown in Fig. 1(a,b).

Experimental studies have thus far primarily focused on developing a mapping between repeating sequence units of spider silk and the basic structural building blocks of fibrils. Two distinct proteins are typically found in dragline silks with similar sequence across species (Gatesy et al. 2001). One of the most studied silk from spiders, *N. Clavipes* dragline silk, contains MaSp1 and MaSp2 proteins, with different repeat units and possibly distinct mechanical functions (Brooks et al. 2005; Hayashi & Lewis 1998; Hayashi et al. 1999; Holland et al. 2008). MaSp1 contains glycine (Gly or G) rich Gly-Gly-X (GGX) repeats with poly-alanine (Ala or A) and GA domains, where X typically stands for alanine, tyrosine, leucine, or glutamine. Whereas MaSp2 also contains poly-Ala domains, it has a repeat unit with high proline content in the form of (GPGQQ/GPGGY). The proline rich segments are intrinsically twisted which inhibit edge-to-edge aggregation of strands, thereby controlling the location and size of beta-sheet nanocrystals in silk. These segments, bearing sequence resemblance to elastin, are thought to form beta-spiral or type II beta-turn structures that provide extensibility through hidden length formation and control the unique thermomechanical properties of spider silk (Gosline et al. 2002; Hayashi et al. 1999; Savage & Gosline 2008). Earlier studies have suggested that MaSp1 is more dominant in the composite morphology of the spider dragline silk than MaSp2, with a ratio of approximately 3:2 or higher, depending on the species (Brooks et al. 2005; Guerette et al. 1996; Hinman & Lewis 1992; Sponner et al. 2005). Recent investigations revealed that anti-parallel beta-sheet crystals at the nanoscale, consisting of highly conserved poly-(Gly[G]-Ala[A]) and poly-Ala repeats found in both commercial and spider silk (Hayashi et al. 1999), play a key role in defining the mechanical properties of silk by providing stiff orderly cross-linking domains embedded in a semi-amorphous matrix with less orderly structures (Lefevre et al. 2007; Thiel et al. 1997; van Beek et al. 2002). Earlier studies have shown that hydration level and solvent conditions such as ion content and pH play a role in the structure and mechanical properties of silk proteins (Dicko et al. 2004; Rammensee et al. 2008). For instance, a unique aspect of silk fibers is their capacity to exhibit a dramatic reduction in length upon hydration; a phenomenon known as supercontraction (Shao & Vollrath 1999; van Beek et al. 1999).

Beta-sheet nanocrystals that employ dense network of hydrogen bonds (Keten & Buehler 2008a; Keten & Buehler 2008b) have dimensions of a few nm and constitute at least 10-15% of the silk volume, while with less orderly extended structures the beta-sheet content can be much higher for most silks (Du et al. 2006; Grubb & Jelinski 1997; Rousseau et al. 2004). The existence of 3_1 helices, and beta-turn or beta-spiral conformations has been suggested for the amorphous domains (Lefevre et al. 2007; Philip et al. 1994; Thiel et al. 1997; van Beek et al. 2002); however, no definite atomistic level structural model has yet been reported. It is anticipated that novel statistical mechanics approaches (Porter et al. 2005), experimental methods, such as x-ray diffraction and scattering (Riekel & Vollrath 2001; Trancik et al. 2006), solid-state NMR (Holland et al. 2008; Jenkins et al. 2010; Simmons et al. 1996; van Beek et al. 1999) and Raman spectroscopy (Lefevre et al. 2007; Rousseau et al. 2004; Rousseau et al. 2009), combined with pioneering multiscale atomistic modeling methods such as those based on density functional theory (Keten et al. 2010; Porter & Vollrath 2008) or molecular dynamics (Brooks 1995; Buehler et al. 2008; Keten & Buehler 2010; Ma & Nussinov 2006) will provide more insight into the atomic resolution structure for complex materials such as spider silk. An earlier study of silk using replica exchange molecular dynamics (Keten & Buehler 2010) has yielded interesting results in comparison with experiments (Holland et al. 2008; Kummerlen et al. 1996; van Beek et al.

2002). However, this study was focused solely for the case of MaSp1 and did not encompass any mechanical characterization. Due to the lack of current large-scale atomic resolution models, the links between genetic makeup, chemical interactions and structure, as well as their associated macroscale mechanical properties remain obscure. Earlier atomistic studies have focused solely on the well-defined crystalline regions consisting of poly-Ala (spider silk) or poly-Gly-Ala repeats (commercial silk) (Keten et al. 2010; Xiao et al. 2009), and studies on the links between structural and nanomechanical features of the nanocomposite structure of silk have remained a challenge.

Here we carry out atomistic simulations to identify nanostructural models of spider silk proteins, with the goal to develop a link between genetic sequence and resulting mechanical properties. The challenges of reaching native (equilibrium) structures within the time-scales accessible to conventional molecular dynamics simulation require enhanced sampling methods such as replica exchange molecular dynamics (REMD) (Sugita & Okamoto 1999). Here we employ REMD to investigate the structures formed by assemblies of short segments of MaSp1 and MaSp2 proteins. Along with other protein structure prediction approaches (Bradley et al. 2005; Zhang 2008), REMD is considered to be an effective tool for investigating folding and aggregation of proteins, as it reduces the likelihood of kinetic trapping at non-native states (Sanbonmatsu & Garcia 2002). Through a fast search of the conformation space at high temperatures and more detailed investigation at low temperatures, it allows the system to overcome energy barriers and local minima corresponding to non-native structures (Feig et al. 2003; Miyashita et al. 2009; Rao & Caflisch 2003; Rhee & Pande 2003) of proteins and allows identifying native protein structures from the amino acid sequence, with atomistic resolution. The overall approach is summarized in Fig. 2(a).

2. MATERIALS AND METHODS

The computational approach used here consists of two steps, (1) structure identification with validation against experimental results, and (2) mechanical analysis of the resulting structures, with subsequent analysis of mechanical properties and deformation mechanisms.

2.1 Structure identification

Primary protein structures (polypeptide sequences) are created with the amino acid sequence of the *N. Clavipes* MaSp1 and MaSp2 proteins, which constitute the majority of the silk's core (Holland et al. 2008). The MaSp1 sequence is (in one-letter amino acid codes): GGAGQGGYGGLGSQGAGRGGLGGQGAG**AAAAAA**GGAGQGGYGGLGSQGAGRGGLGGQG AG. The MaSp2 sequence is: GPGQQGPGGYGPGQQGPGGYGPGQQGPSGPGS**AAAAAAAA**GPGQQGPGGYGPGQQGPGGY GPGQQGPSGPGS.

The monomers consist of two glycine-rich repeating units surrounding a poly-alanine segment to represent a single monocrystal. MaSp1 protein consists primarily of GGX type repeats followed by a GA and poly-Ala region. MaSp2 on the other hand has proline-rich GPGQQGPGGY repeats, and a poly-Ala region. The starting configuration is a lattice structure consisting of anti-parallel arrangement in one direction and parallel arrangement in the other. Previous findings on poly-alanine aggregation suggest that anti-parallel orientation in the hydrogen bonding direction and parallel stacking in the side-chain direction lead to stable beta-sheets (Ma & Nussinov 2002), and hence such an arrangement is considered here to be a good starting point for obtaining assembled structures by silk proteins. In the initial setup, each strand is separated by 10 Å in a square lattice (see Fig. 2(a)). The simulations start from an extended conformation, which is very relevant in the processing of silk where elongational

flow in the spinning duct lead to stretching and alignment of monomers in the concentrated dope, as shown in recent experimental work (Rammensee et al. 2008) (see Fig. 2(b) for a schematic). We note that higher temperature replicas can allow wide sampling around this basic orientation, where strands can rearrange in anti-parallel or parallel fashion as the strands can diffuse within the lattice. At high temperature replicas, weak interactions between the strands can be broken; that is, the system 'melts' and reforms in a wide range of structures that can be investigated in detail at lower temperatures. A key hypothesis here for using an initial orderly lattice arrangement is that edge-to-edge aggregation of strands–similar to amyloids (Kenney et al. 2002)–is the driving force for formation of mono-crystals, rather than local folding of the backbone onto itself through short turn structures. This is supported by recent experiments, which suggest that large extensional forces during spinning, and high concentration of the dope are requirements for crystal formation (Rammensee et al. 2008). This assumption is reasonable in the context of a monocrystal study as pursued here, and does not rule out the possibility of the formation of folded intramolecular contacts at longer length scales. Experimental evidences also suggest that strand orientation in the proximity of the crystals is aligned more or less parallel the fiber axis, also supporting this initial configuration (Rousseau et al. 2009; van Beek et al. 2002).

For structure identification we carry out simulations with Langevin dynamics using CHARMM (BROOKS et al. 1983) and the EEF1.1 force field with a Gaussian effective solvent energy function (Lazaridis & Karplus 1999). The REMD protocol is setup using the MMTSB toolset (Feig et al. 2004). A simulation time step of 2 fs is used by employing the SHAKE algorithm for hydrogen atoms. Solvent friction is added via a Langevin friction term that allows for high mobility and conformational sampling. While the EEF1.1 model has some particular modifications and simplifications on solvent, side-chain and hydrogen bond interactions, it has the benefit of being orders of magnitude faster than other implicit or explicit solvent models, and the solvent volume exclusion model is particularly attractive for large-scale assembly processes as in the case of spider silk. Since force fields are generally parameterized for room temperature calculations, we only pick ensemble structures from the lowest temperature replica and use higher temperatures only for overcoming kinetic trapping and fast conformational search in the REMD scheme.

Initial structures are oriented along the main chain axis using built-in functions in CHARMM (see geometry shown in Fig. 2). We perform long initialization runs to achieve distinct starting configurations to enhance better sampling in the production run. This is followed by a production run starting from the final configurations of the replicas from the initialization run and using an exchange time step of 2 ps to allow the relaxation of the system. Starting from the final configurations of the initial run, we run simulations on 64 replicas distributed over the 300-650 K temperature range (the high temperature replicas ensure that the protein can sample a wide range of conformations, specifically including those that resemble the initial structural ordering of strands). Each replica is simulated for a total of 10 ns, corresponding to a total simulation time of 640 nanoseconds.

An ensemble of structures is analyzed from the last 1,000 exchanges of the production run from the lowest temperature replica, and is used to extract the lowest energy structures (coordinates of all structures generated from our simulations are available from the authors upon request). Representative final structures are selected based on a clustering algorithm based on mutual similarity according to root mean square deviation for each structure. We use the K-means clustering algorithm in the MMTSB toolset (Feig et al. 2004) to identify the largest clusters that have approximately 10% or more presence in the selected set, corresponding to the top 5 clusters. From the selected clusters, we present structures closest to the cluster center as the representative models. Analysis on dihedral angles and alpha-carbon distances are carried out using the complete data set to achieve better statistical

representation. The secondary structure content is calculated in Visual Molecular Dynamics (VMD) (Humphrey et al. 1996) using the STRIDE algorithm.

2.2 Nanomechanical analysis

Mechanical stretching tests are done using a constant loading rate of 2 pN applied every 20 ps, that is, a loading rate of 0.2 N/s. Half of the chains are randomly selected to be pulled in +z direction while the other half is pulled in –z, to impose the characteristic lateral loading of the crystals as experienced in the native silk structure (see schematic shown in Fig. 2(c)).

Force-extension plots are based on the forces applied, and the measured distance between the center of mass of pulled atoms is computed. The analysis on hydrogen bond dynamics is computed using .tcl scripts.

3. RESULTS

Here we present our structural predictions and compare the results with experimental evidence. We focus primarily on torsion angles and secondary structures of most common amino acids, since this approach allows for a direct quantitative comparison that is independent of the size of the system studied. Since the majority of silk consists of glycine and alanine amino acids, we directly compare the dihedral angles of gycine and alanine residues with experimental data on spider silk proteins. For MaSp1 alanine residues (see Fig. 3(a)), we observe that the most common phi-psi angle value is around (-150,135) in excellent agreement with experimental findings that suggested (-135,150), corresponding to a beta-sheet structure. For glycine (see Fig. 3(b)), our results show a symmetry around the origin, where we observe a wide distribution around approximately (+/-75,-/+75), in agreement with experimental findings around (+/-60,-/+135)(van Beek et al. 2002) that also have symmetry. Similar results are observed for MaSp2 (Fig. 3(c-d)), however, a wider distribution of glycine dihedral angles is evident from the Ramachandran plot.

While mapping of the poly-alanine regions to beta-sheet conformation is straightforward from the data on alanine residues for both structures, the glycine regions require more data to be able to distinguish them within common protein structures. A close look at proline residues in MaSp2 shows peaks around (-60,-30) and (-60,120), corresponding to type I and type II beta-turns that would incorporate adjacent Gly residues. From the ensemble of structures obtained from MaSp1 replica exchange simulations, we consider the alpha-carbon atom distances between glycine residues that are three residues apart, to characterize the basic repeat unit length of the GGX domains. The probability distribution of Cα(i)-Cα(i+3) distances measured from atomic coordinates has a distinct peak around 9 Å. This value is greater than most known beta-turn structures and falls short of the anti-parallel beta-sheet conformations that typically exceed 10 Å. These domains have interchain hydrogen bonding that is less oriented and sparser. Based on these observations, we conclude that these disorderly structures in MaSp1 resemble the characteristics of 3_1-helices, which are the most likely conformations taken by Gly-rich domains, supporting earlier hypotheses reported in the literature (Kummerlen et al. 1996; van Beek et al. 2002). The density of hydrogen bonding in these regions is significantly reduced compared with the orderly network in beta-sheet nanocrystals, suggesting that moderating the number and orientation of intermolecular interactions leads to the characteristic heterogeneous structural arrangement observed in silks.

For illustrative purposes, a representative collection of selected structures (see Methods) obtained from the simulations is shown in Fig. 4 where percentages of different secondary structures are illustrated in each sub-panel (I-V). For both MaSp1 and MaSp2, a significant percentage of residues are found to be

in semi-extended, disorderly conformations in agreement with the discussion put forth based on experimental studies (Rammensee et al. 2008). Our results suggest that poly-alanine regions have an extremely high propensity for aggregating into crystalline beta-sheet structures by inter-strand hydrogen bond formation and inter-sheet stacking in the side-chain direction. Along with glycine regions that also form extended regions, the beta-sheet percentage ranges from 28% to 55% for the sequences studied in this work. These results compare well with recent NMR studies on dragline silk, which have indicated 34% beta-sheet content (Jenkins et al. 2010). This highly orderly domain is dispersed within the glycine-rich repeat units, which are still fairly oriented but much less-orderly, forming more amorphous structures. A key finding from the secondary structure content analysis is the lack of any alpha-helix conformation of the MaSp1 and MaSp2 silk constituents, supporting a wide range of experimental evidence that ruled out this conformation (Holland et al. 2008; Jelinski 1998; Jenkins et al. 2010; van Beek et al. 2002). Instead, our results suggest that disorderly structures resembling 3_1-helices and beta-turns dominate non beta-sheet conformations in these proteins. The higher content of proline in MaSp2 leads to more disorderly structures in amorphous regions and well-defined beta-sheet crystal regions. This is evident from both the lower ratio of beta-sheets in this sequence, and also the shorter lengths of beta-strands when comparing MaSp1 structures in Fig. 4 (a) with those of MaSp2 in Fig. 4 (b).

Altogether, the structure identification and comparison with available experimental results suggest that the models obtained from our simulations resemble those found in native spider silk. Overall, crystal structures consisting of poly-Ala repeats have the size of 2-3 nanometers in the chain direction, with partial beta-sheet domains in semi-crystalline regions, particularly for MaSp1. Stacks of 2 to 4 sheets (1-2 nm) and up to 8 strands per sheet (2-4 nm) are observed in simulations (but these values may increase if the number of chains considered is increased). The amorphous domains are semi-extended, and the average lateral length per residue for both structures emerges to be around 2 to 3 nm, which is shorter than fully extended structures that would require more than 3 nm length. We do not observe larger folds where sheets are formed by polypeptide strands with self-interactions, but such morphologies cannot be ruled out considering that actual spider silk sequences are much longer and therefore flexible than the short segments we have studied here. In this regard, this study sets the stage now to explore the mechanical properties of both the MaSp1 and MaSp2 structures using ultra-large-scale molecular dynamics simulations.

We proceed with carrying out constant loading rate mechanical shear simulations on the selected structures, where randomly selected strands are pulled in opposite directions to mimic the relevant lateral (shear) loading conditions of small crystals in spider silk. The resulting force-displacement curves from these simulations are shown in Fig. 5(a-b). Both MaSp1 and MaSp2 show a characteristic smooth curve with three regimes, where the relative moduli of these regimes depend on the secondary structure content. For molecules with high turn ratio and low beta-sheet content (see MaSp1-IV and MaSp2-II), we observe an initial stiff regime, followed by a softer regime, followed by a very stiff regime leading to failure. On the other hand, systems with very high beta-sheet content (MaSp1-I, MaSp2-I) show a monotonically stiffening force-extension response. This suggests that the characteristic yielding behavior at the molecular level of the hierarchy of silk is controlled by the ratio of turn to beta-sheet structures, where a higher turn ratio leads to the emergence of this phenomenon. This is an alternative means of achieving high initial stiffness in comparison with for instance the MaSp1-I structure, which has more beta-sheet content and therefore exhibits high initial stiffness as well, but provides much less extensibility. We find that MaSp1 generally exhibits a larger deviation in the initial end-to-end length, owing to the wider secondary structure distribution. The effect of this is evident from the analysis displayed in Fig. 5(c), which illustrates the yielding behavior of both structures and much larger variation in response in MaSp1, where increasing the turn content and inter-

chain hydrogen bonding increases the initial stiffness and extensibility of the assembly, thereby improving the toughness as well. The failure strength of MaSp1 and MaSp2 seem to be similar in our simulations, as shown in Fig. 5(d).

The force extension curves obtained here also show minor deviations from the inextensible chain models commonly used for polymer materials. At low forces, this is due to the rearrangement and rupture of H-bonds in the amorphous domains. At high force, the crystal morphology can change due to large shear stresses as well as transverse compression, where the latter causes buckling and collapse of the sheet into a smaller, more compact formation.

Some of the structural transformations and failure mechanisms can be observed from the trajectories obtained from stretching simulations of the spider silk assemblies. Fig. 6(a-b),(i-v) show snapshots of deformation from MaSp1 and MaSp2 stretching simulations. As evident from the snapshots and the analysis presented in Fig. 7(a-b), amorphous domains stretch significantly with applied force, and a transition from turn to beta-sheet structures can be observed for both MaSp1 and MaSp2 (regime just before the point marked by a red arrow in Figure 7). The percent change seems to be larger for MaSp2, since in their unstretched configuration proline residues act as beta-sheet breakers. A key observation is that failure of the system happens by sliding of strands with respect to each other, which can occur only upon breaking of the hydrogen bonds in the crystalline domain. In our simulations, failure occurs at the interface region with solvent at the boundary of the crystal, leaving part of the crystal intact even after failure.

Representative plots from MaSp1 and MaSp2 structures with highest turn content are shown in Fig. 7. A comparison of panel (a) and panel (b) shows the strain softening behavior that occurs when the turn/beta-sheet ratio is high in both structures. Upon initial yielding, number of hydrogen bonds in the system and also the turn content decreases, as shown in Fig. 7 (a) and (b) respectively for Masp1 and MaSp2. Further extension leads to a slight increase in the number of H-bonds and formation of beta-sheet structures in the amorphous region. This is followed by a final stiff regime dominated by the stretching of covalent bonds, which ultimately leads to rapid rupture of many hydrogen bonds (shown with a red arrow) in the non-crystalline domains. Eventually, a final stiff regime is reached where covalent bonds in the peptide backbone are stretched out (regime after the point marked by a red arrow in Figure 7), which initiates rapid rupture of many hydrogen bonds in the amorphous domains. The system fails upon breaking of hydrogen bonds in the crystal, and the subsequent sliding of strands. The constitutive sigmoidal behavior of the force-displacement graphs observed here for both structures agrees qualitatively with the macroscale response of spider silk.

4. DISCUSSIONS AND CONCLUSION

We have presented results from atomistic simulations on MaSp1 and MaSp2 protein segments of the spider dragline silk from *N. Clavipes*. We have illustrated that the REMD simulation method establishes an atomistic basis for a wide variety of findings from experimental approaches of studies of the silk nanostructure. Key finding from secondary structure and dihedral angle analysis have shown that poly-Ala regions in MaSp1 and MaSp2 form beta-sheet crystals, whereas the glycine rich regions form semi-extended 3_1 helix type structures as well as beta-turns (Fig. 3). No evidence for alpha-helix or beta-helix structure formation is observed in our simulations for both MaSp1 and MaSp2.

Our results confirm that MaSp1 and MaSp2 form different nanostructures. MaSp1 tends to form more orderly extended structures with beta-sheet domains, whereas turn structures and a greater level of amorphousness are observed in MaSp2, due to the natural twist of the proline segments, which limits

the crystalline fraction of the protein assemblies (Figs. 3 and 4). The existence of semi-extended domains in the so-called amorphous matrix may be the molecular source of the large semi-crystalline fraction observed in silks, and also form the basis of the so-called "pre-stretched" molecular configuration. The presence of less-dense hydrogen bonding in glycine-rich regions compared with alanine-rich regions indicates at least a two-phase system with complementary mechanical functions: extensibility and fracture strength. This is more evident in MaSp2, which has apparent features resulting from the proline rich sequence, such as higher disorder and well-defined crystal regions. Of particular importance for this effect is the lack of amide bonds and the torsional twisting at proline sites, which controls morphology development as the proline residues have varying capacity to form hydrogen bonding at different extension states. In the actual morphology of dragline silk, both proteins can be observed at varying amounts depending on the species (Brooks et al. 2005; Guerette et al. 1996; Hinman & Lewis 1992; Sponner et al. 2005). The current study focused on isolated systems of MaSp1 and MaSp2, but the same simulation protocol can be applied to understand the structure and behavior of mixed systems, and the possible role of proline in controlling intermolecular contacts between MaSp1 and MaSp2. Overall, the results reported here are strongly supportive of experimental findings on the spider dragline silk that suggest similar structural characteristics, in particular, the existence of a Gly-rich amorphous phase and Ala-rich crystalline phase (Thiel et al. 1997; van Beek et al. 2002).

The mechanical stretching simulations hint towards a sigmoidal constitutive behavior as observed in the macroscale response of spider silk, as shown in Fig. 5(c). Our results provide insight into the nanoscale deformation mechanisms, as shown in the simulation snapshots in Fig. 6 and the detailed analysis in Fig. 7. We find that the initial softening of the system is governed by hydrogen bond breaking in the amorphous domain. Strain stiffening is observed once the covalent chains are highly extended after rupture of a large number of hydrogen bonds in the non-crystalline domains. The eventual failure of the system occurs when strands begin to slide across each other due to severed hydrogen bonds and side-chain interactions in the crystal. Considering that the initial end to end-length of the systems studied range from 8 nm to 14 nm and failure occurs at around 25 nm or more, the failure strains observed here seem to be much larger than experimental studies that show failure strains around 40% (Du et al. 2006). The source of this discrepancy may be several factors in our simulations that differ from experiments. First, the results for strain will depend on sequence; here we study two amorphous domains and one crystal domain, hence our system is likely to be more extensible. The second reason is that in the actual morphology of silk, there will be heterogeneity and defects, and orientation of the forces with respect to chains will not be perfect. Hence the extensibility of the chains will be further reduced, as they will slide across each other more easily in non-uniform loading. A third possible reason is that current simulations correspond to hydrated systems with no control on prestretching. Prestretching of the strands during silk synthesis may contribute to the lower failure strain of the actual systems. More analysis and upscaling of our results will have to be done to make a rigorous link between out nanoscale results and macroscale experiments. The failure strength of the molecular assembly shows less variation between sequences; rather, it is observed to depend more on the morphology of the crystal and the amorphous structures. Based on our simulations, we note that mechanical stretching of the strands may play a role in controlling crystal size, where compression in the transverse direction causes collapse of the crystals into a more compact formation. This may be an important consideration for the influence of mechanical strain on the morphology of the silk crystals (Grubb & Jelinski 1997).

In summary, we have developed and applied a framework for predicting the nanostructure of spider silk using atomistic principles that could be widely applicable not only to different types of silks but could also to other biopolymers, and thereby provide a link between genetics and material properties. This materiomics framework provides a powerful tool to explore fundamental structure-property

relationships of complex biological materials such as spider silk. Our atomistic calculations resolve some of the controversies regarding the structure of the amorphous domains in silk, by shedding light on the semi-extended, well-oriented and more sparsely hydrogen bonded structures in amorphous domains. To the best of our knowledge, the results reported here – combining REMD structure identification with experimental validation and a subsequent detailed nanomechanical analysis with a study of fundamental deformation mechanisms – is the first of its kind for spider silk, and only the beginning. We expect that future studies could be focused on improved structure prediction and a wide variation of conditions during mechanical analysis.

Our findings may set the stage for more extensive full-atomistic mechanics studies on silk domains that will contribute towards an improved understanding of the source of the strength and toughness of this biological superfiber, where the structures identified here could be subjected to mechanical loading under varying conditions, such as changes in solvent or pH, and even under variations of the amino acid sequence of the constituting protein domains. Future investigations could also focus on much larger, mixed MaSp1/MaSp2 systems in a variety of conditions or on the impact of mechanical constraints on the assembly and initial structure formation.

Acknowledgements: This work was supported by the Office of Naval Research (N00014-08-1-00844 and N00014-10-1-0562). This research was supported by an allocation of advanced computing resources supported by the National Science Foundation (TeraGrid, grant # TG-MSS080030). Fruitful discussions of our results with David Kaplan are greatly appreciated. We thank Ross Walker and Michael Feig for helpful discussions on the replica exchange simulation protocol.

References

Becker, N., Oroudjev, E., Mutz, S., Cleveland, J. P., Hansma, P. K., Hayashi, C. Y., Makarov, D. E. & Hansma, H. G. 2003 Molecular nanosprings in spider capture-silk threads. *Nature Materials* **2**, 278-283.

Bradley, P., Misura, K. M. S. & Baker, D. 2005 Toward high-resolution de novo structure prediction for small proteins. *Science* **309**, 1868-1871.

Brooks, A. E., Steinkraus, H. B., Nelson, S. R. & Lewis, R. V. 2005 An investigation of the divergence of major ampullate silk fibers from Nephila clavipes and Argiope aurantia. *Biomacromolecules* **6**, 3095-3099.

BROOKS, B., BRUCCOLERI, R., OLAFSON, B., STATES, D., SWAMINATHAN, S. & KARPLUS, M. 1983 CHARMM - A PROGRAM FOR MACROMOLECULAR ENERGY, MINIMIZATION, AND DYNAMICS CALCULATIONS. *JOURNAL OF COMPUTATIONAL CHEMISTRY* **4**, 187-217.

Brooks, C. L. 1995 Methodological Advances in Molecular-Dynamics Simulations of Biological-Systems. *Current Opinion in Structural Biology* **5**, 211-215.

Buehler, M. J., Keten, S. & Ackbarow, T. 2008 Theoretical and computational hierarchical nanomechanics of protein materials: Deformation and fracture. *Progress in Materials Science* **53**, 1101-1241.

Buehler, M. J. & Yung, Y. C. 2009 Deformation and failure of protein materials in physiologically extreme conditions and disease. *Nature Materials* **8**, 175-188.

Dicko, C., Vollrath, F. & Kenney, J. M. 2004 Spider silk protein refolding is controlled by changing pH. *Biomacromolecules* **5**, 704-710.

Du, N., Liu, X. Y., Narayanan, J., Li, L. A., Lim, M. L. M. & Li, D. Q. 2006 Design of superior spider silk: From nanostructure to mechanical properties. *Biophysical journal* **91**, 4528-4535.

Feig, M., Karanicolas, J. & Brooks, C. L. 2004 MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *Journal of Molecular Graphics & Modelling* **22**, 377-395.

Feig, M., MacKerell, A. D. & Brooks, C. L. 2003 Force field influence on the observation of pi-helical protein structures in molecular dynamics simulations. *Journal of Physical Chemistry B* **107**, 2831-2836.

Fratzl, P. & Weinkamer, R. 2007 Nature's hierarchical materials. *Progress in Materials Science* **52**, 1263-1334.

Gatesy, J., Hayashi, C., Motriuk, D., Woods, J. & Lewis, R. 2001 Extreme diversity, conservation, and convergence of spider silk fibroin sequences. *Science* **291**, 2603-2605.

Gosline, J., Lillie, M., Carrington, E., Guerette, P., Ortlepp, C. & Savage, K. 2002 Elastic proteins: biological roles and mechanical properties. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **357**, 121-132.

Grubb, D. T. & Jelinski, L. W. 1997 Fiber morphology of spider silk: The effects of tensile deformation. *Macromolecules* **30**, 2860-2867.

Guerette, P. A., Ginzinger, D. G., Weber, B. H. F. & Gosline, J. M. 1996 Silk properties determined by gland-specific expression of a spider fibroin gene family. *Science* **272**, 112-115.

Hayashi, C. Y. & Lewis, R. V. 1998 Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks. *Journal of Molecular Biology* **275**, 773-784.

Hayashi, C. Y., Shipley, N. H. & Lewis, R. V. 1999 Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. *International Journal of Biological Macromolecules* **24**, 271-275.

Hinman, M. B. & Lewis, R. V. 1992 Isolation of a Clone Encoding a 2nd Dragline Silk Fibroin - Nephila-Clavipes Dragline Silk Is a 2-Protein Fiber. *Journal of Biological Chemistry* **267**, 19320-19324.

Holland, G. P., Creager, M. S., Jenkins, J. E., Lewis, R. V. & Yarger, J. L. 2008 Determining secondary structure in spider dragline silk by carbon-carbon correlation solid-state NMR spectroscopy. *Journal of the American Chemical Society* **130**, 9871-9877.

Humphrey, W., Dalke, A. & Schulten, K. 1996 VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **14**, 33-*&*.

Jelinski, L. W. 1998 Establishing the relationship between structure and mechanical function in silks. *Current Opinion in Solid State & Materials Science* **3**, 237-245.

Jenkins, J. E., Creager, M. S., Lewis, R. V., Holland, G. P. & Yarger, J. L. 2010 Quantitative Correlation between the Protein Primary Sequences and Secondary Structures in Spider Dragline Silks. *Biomacromolecules* **11**, 192-200.

Kenney, J. M., Knight, D., Wise, M. J. & Vollrath, F. 2002 Amyloidogenic nature of spider silk. *European Journal of Biochemistry* **269**, 4159-4163.

Keten, S. & Buehler, M. J. 2008a Asymptotic strength limit of hydrogen bond assemblies in proteins at vanishing pulling rates. *Physical Review Letters*.

Keten, S. & Buehler, M. J. 2008b Geometric Confinement Governs the Rupture Strength of H-bond Assemblies at a Critical Length Scale. *Nano Lett.* **8**, 743-748.

Keten, S. & Buehler, M. J. 2010 Atomistic model of the spider silk nanostructure. *Applied Physics Letters*.

Keten, S., Xu, Z., Ihle, B. & Buehler, M. J. 2010 Nanoconfinement controls stiffness, strength and mechanical toughness of beta-sheet crystals in silk. *in submission*.

Kummerlen, J., vanBeek, J. D., Vollrath, F. & Meier, B. H. 1996 Local structure in spider dragline silk investigated by two-dimensional spin-diffusion nuclear magnetic resonance. *Macromolecules* **29**, 2920-2928.

Lazaridis, T. & Karplus, M. 1999 Effective energy function for proteins in solution. *Proteins-Structure Function and Genetics* **35**, 133-152.
LeDuc, P. R. & Robinson, D. N. 2007 Using Lessons from Cellular and Molecular Structures for Future Materials. *Advanced Materials* **19**, 3761-3770.
Lefevre, T., Rousseau, M. E. & Pezolet, M. 2007 Protein secondary structure and orientation in silk as revealed by Raman spectromicroscopy. *Biophysical Journal* **92**, 2885-2895.
Ma, B. Y. & Nussinov, R. 2002 Molecular dynamics simulations of alanine rich beta-sheet oligomers: Insight into amyloid formation. *Protein Science* **11**, 2335-2350.
Ma, B. Y. & Nussinov, R. 2006 Simulations as analytical tools to understand protein aggregation and predict amyloid conformation. *Current Opinion in Chemical Biology* **10**, 445-452.
Miyashita, N., Straub, J. E., Thirumalai, D. & Sugita, Y. 2009 Transmembrane Structures of Amyloid Precursor Protein Dimer Predicted by Replica-Exchange Molecular Dynamics Simulations. *Journal of the American Chemical Society* **131**, 3438-+.
Philip, M. C., Stephen, A. F., Margaret, A. A., John, W. S., David, L. K., Adams, W. W., Ronald, K. E., David, M. & Deborah, L. V. 1994 Mechanical and thermal properties of dragline silk from the spider <I>Nephila clavipes</I>. *Polymers for Advanced Technologies* **5**, 401-410.
Porter, D. & Vollrath, F. 2008 The role of kinetics of water and amide bonding in protein stability. *Soft Matter* **4**, 328-336.
Porter, D., Vollrath, F. & Shao, Z. 2005 Predicting the mechanical properties of spider silk as a model nanostructured polymer. *European Physical Journal E* **16**, 199-206.
Rammensee, S., Slotta, U., Scheibel, T. & Bausch, A. R. 2008 Assembly mechanism of recombinant spider silk proteins. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 6590-6595.
Rao, F. & Caflisch, A. 2003 Replica exchange molecular dynamics simulations of reversible folding. *Journal of Chemical Physics* **119**, 4035-4042.
Rhee, Y. M. & Pande, V. S. 2003 Multiplexed-replica exchange molecular dynamics method for protein folding simulation. *Biophysical Journal* **84**, 775-786.
Riekel, C. & Vollrath, F. 2001 Spider silk fibre extrusion: combined wide- and small-angle X-ray microdiffraction experiments. *International Journal of Biological Macromolecules* **29**, 203-210.
Rousseau, M. E., Lefevre, T., Beaulieu, L., Asakura, T. & Pezolet, M. 2004 Study of protein conformation and orientation in silkworm and spider silk fibers using Raman microspectroscopy. *Biomacromolecules* **5**, 2247-2257.
Rousseau, M. E., Lefevre, T. & Pezolet, M. 2009 Conformation and Orientation of Proteins in Various Types of Silk Fibers Produced by Nephila clavipes Spiders. *Biomacromolecules* **10**, 2945-2953.
Sanbonmatsu, K. Y. & Garcia, A. E. 2002 Structure of Met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. *Proteins-Structure Function and Genetics* **46**, 225-234.
Savage, K. N. & Gosline, J. M. 2008 The role of proline in the elastic mechanism of hydrated spider silks. *Journal of Experimental Biology* **211**, 1948-1957.
Shao, Z. Z. & Vollrath, F. 1999 The effect of solvents on the contraction and mechanical properties of spider silk. *Polymer* **40**, 1799-1806.
Shao, Z. Z. & Vollrath, F. 2002 Materials: Surprising strength of silkworm silk. *Nature* **418**, 741-741.
Simmons, A. H., Michal, C. A. & Jelinski, L. W. 1996 Molecular orientation and two-component nature of the crystalline fraction of spider dragline silk. *Science* **271**, 84-87.
Sponner, A., Schlott, B., Vollrath, F., Unger, E., Grosse, F. & Weisshart, K. 2005 Characterization of the protein components of Nephila clavipes dragline silk. *Biochemistry* **44**, 4727-4736.

Sugita, Y. & Okamoto, Y. 1999 Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **314**, 141-151.
Termonia, Y. 1994 Molecular Modeling of Spider Silk Elasticity. *Macromolecules* **27**, 7378-7381.
Thiel, B. L., Guess, K. B. & Viney, C. 1997 Non-periodic lattice crystals in the hierarchical microstructure of spider (major ampullate) silk. *Biopolymers* **41**, 703-719.
Trancik, J. E., Czernuszka, J. T., Bell, F. I. & Viney, C. 2006 Nanostructural features of a spider dragline silk as revealed by electron and X-ray diffraction studies. *Polymer* **47**, 5633-5642.
van Beek, J. D., Hess, S., Vollrath, F. & Meier, B. H. 2002 The molecular structure of spider dragline silk: Folding and orientation of the protein backbone. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 10266-10271.
van Beek, J. D., Kummerlen, J., Vollrath, F. & Meier, B. H. 1999 Supercontracted spider dragline silk: a solid-state NMR study of the local structure. *International Journal of Biological Macromolecules* **24**, 173-178.
Vollrath, F. & Knight, D. P. 2001 Liquid crystalline spinning of spider silk. *Nature* **410**, 541-548.
Xiao, S. B., Stacklies, W., Cetinkaya, M., Markert, B. & Grater, F. 2009 Mechanical Response of Silk Crystalline Units from Force-Distribution Analysis. *Biophysical Journal* **96**, 3997-4005.
Zhang, Y. 2008 Progress and challenges in protein structure prediction. *Current Opinion in Structural Biology* **18**, 342-348.

Short title for page headings: Structure and mechanics of spider silk

Figure captions

FIG. 1. Hierarchical structure of spider silk (panel (a)). The work reported here is focused on the scale of the nanocomposite structure, where beta-sheet nanocrystals are immersed in a matrix of semi-amorphous protein (panel (b)). The focus of the present study is on a single beta-sheet nanocrystal embedded in the semi-amorphous matrix.

FIG. 2. Simulation protocol and representative structure results. Panel (a) summarizes the approach taken here to identify the nanostructure of spider silk proteins, here focused on the MaSp1 and MaSp2 silk sequences from the *N. Clavipes* spider. Monomers representing sections of the MaSp1 and MaSp2 proteins (containing a poly-alanine repeat in the center) are used as the basic building block. Replica exchange simulations are carried out at multiple temperatures, and an ensemble of most likely, final structureas are compared with experimental evidence. Panel (b) illustrates the natural process of silk assembly (and fiber formation) during which silk proteins are subject to shear. The natural process of shearing and alignment of protein monomers motivates our choice of the initial geometry shown in panel (a). Panel (c) shows the mechanical loading condition employed in our simulations.

FIG. 3. Comparison of resulting structures for MaSp1 and MaSp2 with experimental data for validation. Panels (a-d): For alanine residues, we observe that the most common phi-psi angle value is around (-150,135) in excellent agreement with experimental findings that suggested (-135,150) that correspond to a beta-sheet structure. For glycine residues, we observe a wide distribution around approximately (+/-75,-/+75) with symmetry around the origin, in line with experimental findings around (+/60,-/+135) that also show symmetric distribution. The wider distributions in MaSp2 may be due to a more amorphous structure caused by high proline content. Panel (e): Dihedral angle distribution of proline residues cluster around (-60,-30) and (-60,120), which correspond to type I and type II beta-turn conformations respectively.

FIG. 4. Secondary structure distribution of selected structures. We select the most representative structures from the lowest temperature replica, and calculate the secondary structure distribution for MaSp1 (panel (a)) and MaSp2 (panel (b)). Subpanels I-V indicate the five most likely structures selected for each sequence. The coloring is based on structural configuration, where yellow represents beta-sheet and extended structures. Panel (a) illustrates selected structures for MaSp1. The majority of the structures are observed to be in beta-sheet or beta-turn conformation, and for MaSp1, beta-sheet content is higher than MaSp2 (panel (b)) due to the lack of proline residues that reduce chain aggregation in sheet form. The relative content of secondary structure controls the mechanical properties of protein materials; where greater crystallinity typically means greater strength, whereas turn structures provide hidden length required for extensibility and toughness. Panel (b) illustrates representative structures for MaSp2. Insets illustrate the stacking formation of the beta-sheets in the poly-Ala regions. The results consistently illustrate that poly-Ala regions form highly orderly beta-sheet crystals whereas the glycine rich repeat units are less-orderly, forming more amorphous domains.

FIG. 5. Force-displacement curves for selected structures. Figure shown illustrates the response of selected structures to shear forces applied to alternating strands. Force values shown are force applied per polypeptide strand. Panels (a-b) illustrate curves obtained from MaSp1 and MaSp2. The forces cause tensile stretching of the strands, where a strain stiffening behavior is evident once the chain reaches a certain length, independent of the chain's initial stretch state. The responses are similar for MaSp1 and MaSp2, however, it depends on the secondary structure content of the system. As shown in

panel (c), MaSp1 structures have a large variation on beta-sheet vs. turn content, which leads to distinctly different mechanical responses. Solid lines shown cases having the largest turn content, whereas dashed lines illustrates structures with more beta-sheets. As the turn ratio increases, an initial stiff regime is observed followed by softening, followed by stiff bond stretching regime. For extended structures, the initial stiff regime disappears, and the typical strain stiffening behavior of polymer chains can be observed. The source of this difference is the existence of denser hydrogen bonding in amorphous regions due to turn formation, which leads to higher stiffness and energy dissipation for structures containing more turns. Lower variation of turn and beta-sheet content in MaSp2 leads to the reduced variation of the mechanical response for this structure. As can be inferred from panel (d), the failure strength of both structures is similar. Since the failure strength is controlled largely by the crystals, which have almost identical sequence in both cases, the morphology of the crystal plays most likely the key role in determining the failure strength.

FIG. 6. Stretching and structural transformation of the proteins. The figure illustrates the stretching behavior of the amorphous domains and crystals under shear forces for MaSp1 (panel (a)) and MaSp2 (panel (b)). Panels (i-v) illustrate the time sequence of events during the stretching simulations. As evident from the time sequence of snapshots (i-v, notation different here to differentiate time sequence from the structure prediction panels shown in Fig. 4), amorphous domains stretch significantly, and a transition from turn to beta-sheet structures are observed. The GGX repeats in MaSp1 are capable of forming beta-sheets during stretching, whereas this is observed to a lesser extent in MaSp2, in particular at low extension. A key observation is that failure of the system happens by sliding of strands with respect to each other upon breaking of the hydrogen bonds and side-chain contacts in the crystalline domain. This typically occurs at the interface region with solvent at the boundary of the crystal, leaving part of the crystal intact even after failure of the structure.

FIG. 7. Representative plots are shown from MaSp1 and MaSp2 structures with the highest turn content. Comparison of panel (a) and panel (b) shows the strain softening behavior that occurs when the turn/beta-sheet ratio is high in the initial structures. Upon initial yielding, the number of hydrogen bonds in the system and also the turn content decreases, however, further stretching leads to a slight increase in the number of H-bonds and formation of beta-sheet structures in the amorphous region. This is followed by a final stiff regime dominated by the stretching of covalent bonds, which ultimately leads to rapid rupture of many hydrogen bonds (shown with a red arrow) in the non-crystalline domains. The system fails upon breaking of hydrogen bonds in the crystal, and sliding of beta-strands. The characteristic sigmoidal force-extension behavior observed here for both structures shows resemblance to the macroscale response of spider silk.

Figures

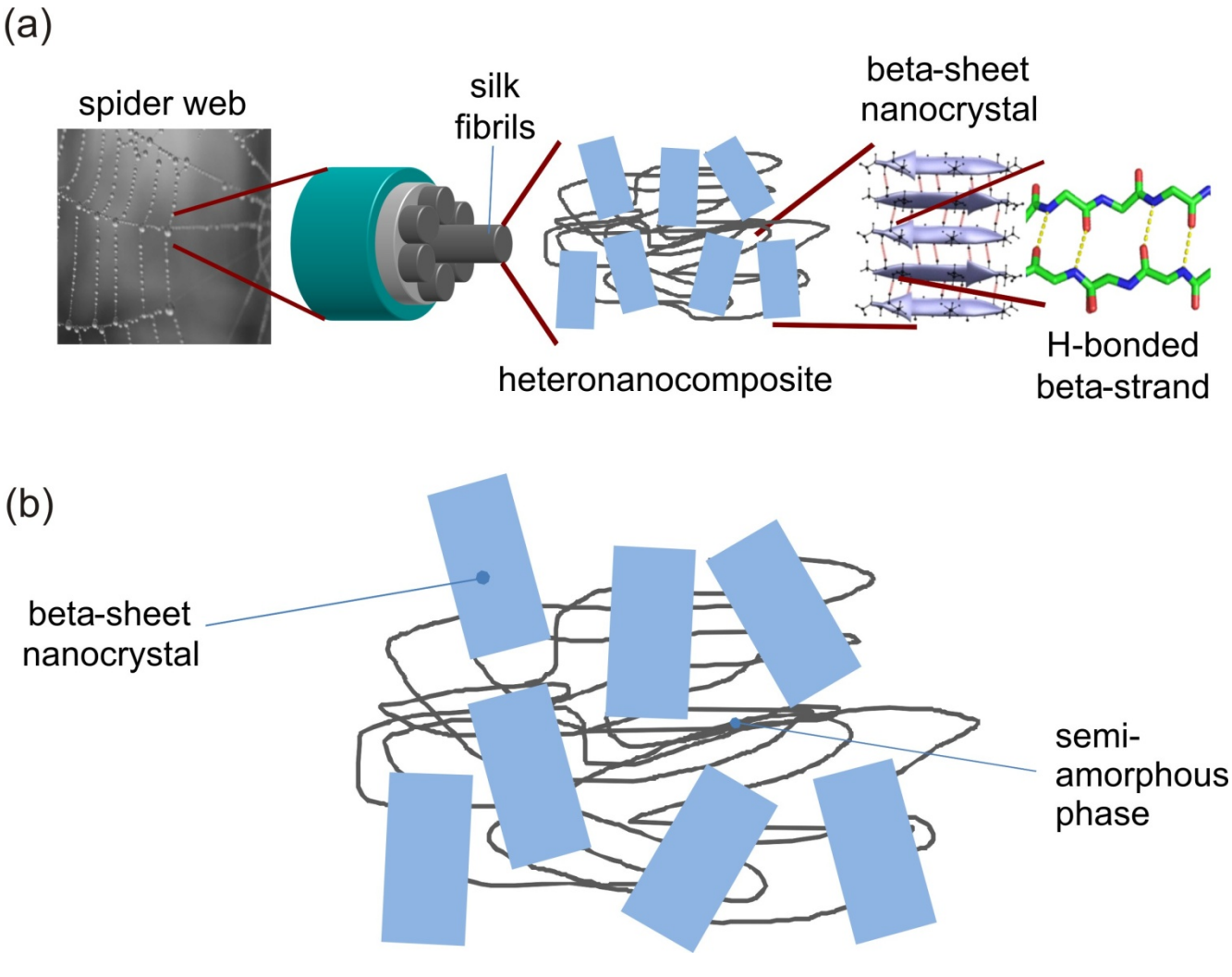

Figure 1

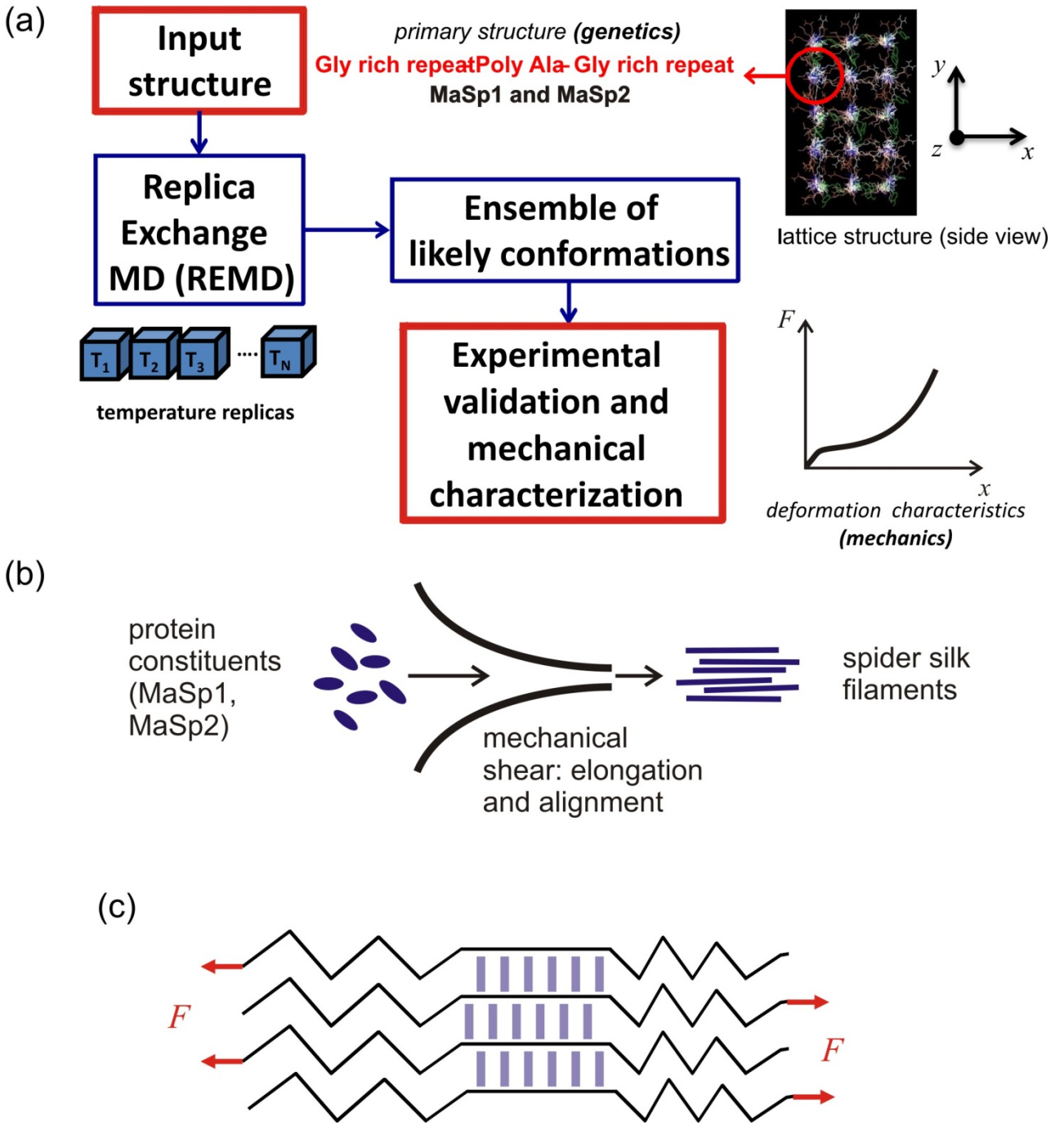
(a)
Input structure
primary structure (genetics)
Gly rich repeat Poly Ala Gly rich repeat
MaSp1 and MaSp2
y
z
x
lattice structure (side view)
Replica Exchange MD (REMD)
Ensemble of likely conformations
T1 T2 T3 .... TN
temperature replicas
Experimental validation and mechanical characterization
F
x
deformation characteristics
(mechanics)
(b)
protein constituents (MaSp1, MaSp2)
mechanical shear: elongation and alignment
spider silk filaments
(c)
F
F

Figure 2

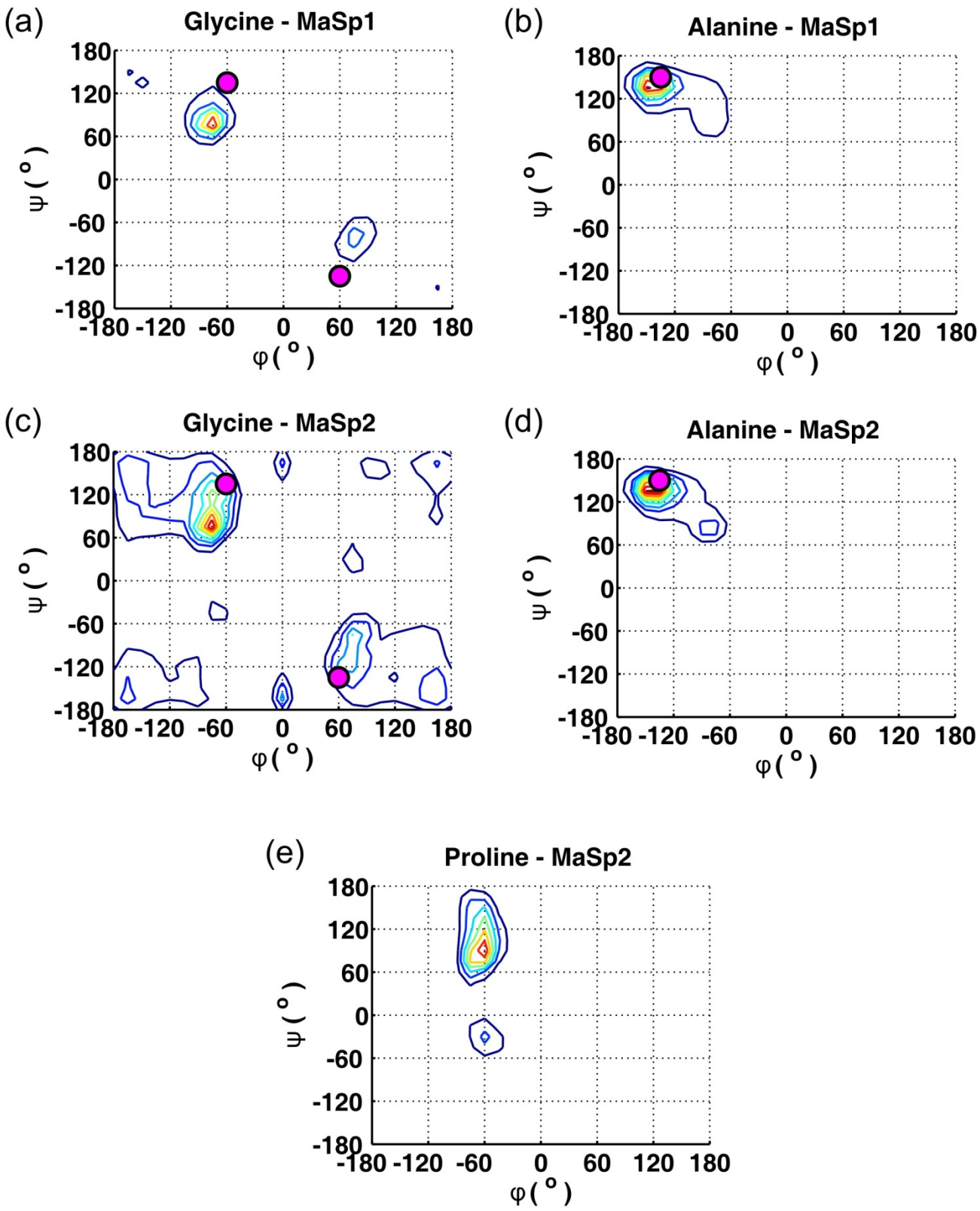
(a)
Glycine - MaSp1
(b)
Alanine - MaSp1
(c)
Glycine - MaSp2
(d)
Alanine - MaSp2
(e)
Proline - MaSp2
ψ(°)
φ(°)
180
120
60
0
-60
-120
-180

Figure 3

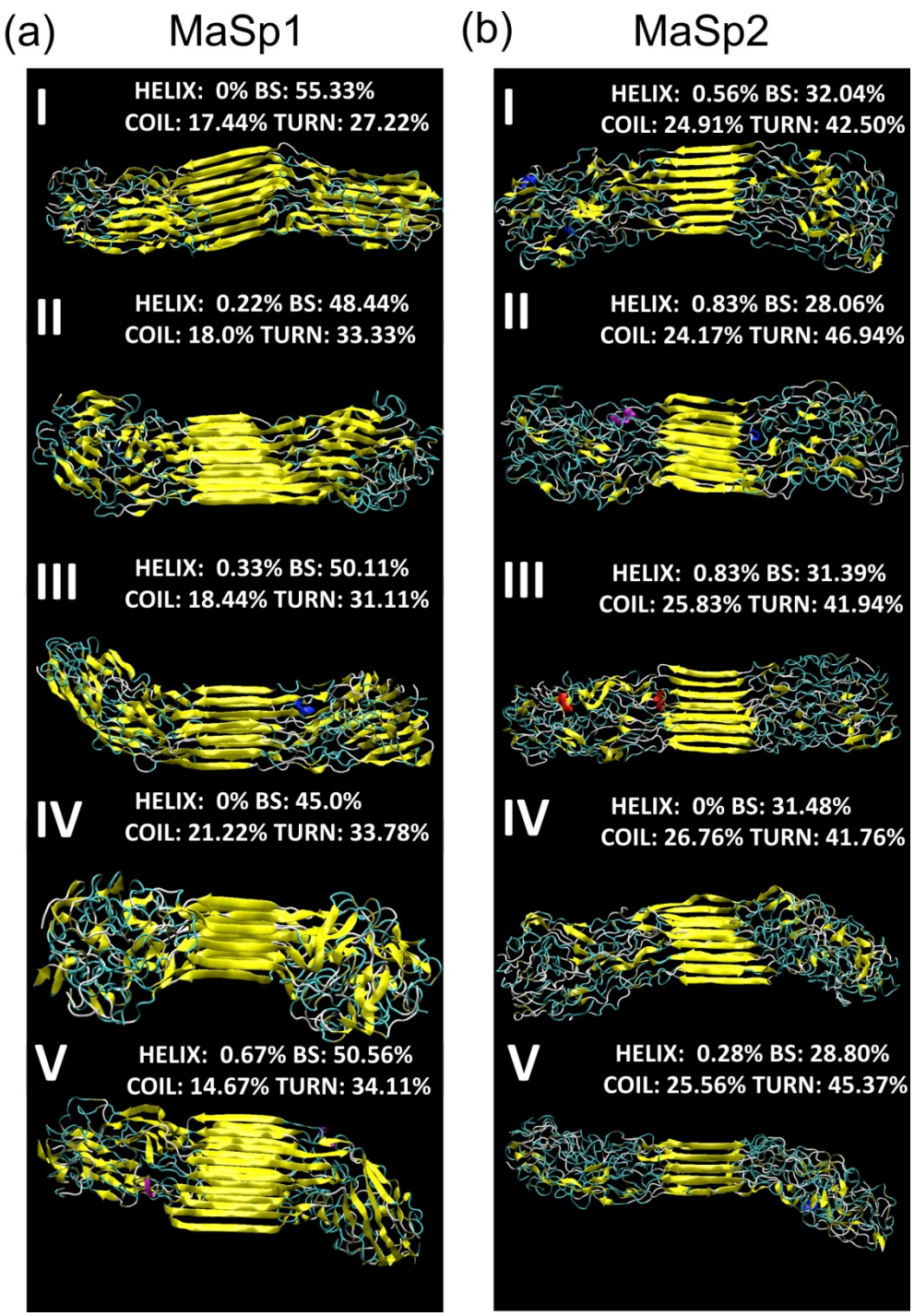
(a) MaSp1
(b) MaSp2
I
HELIX: 0% BS: 55.33%
COIL: 17.44% TURN: 27.22%
II
HELIX: 0.22% BS: 48.44%
COIL: 18.0% TURN: 33.33%
III
HELIX: 0.33% BS: 50.11%
COIL: 18.44% TURN: 31.11%
IV
HELIX: 0% BS: 45.0%
COIL: 21.22% TURN: 33.78%
V
HELIX: 0.67% BS: 50.56%
COIL: 14.67% TURN: 34.11%
I
HELIX: 0.56% BS: 32.04%
COIL: 24.91% TURN: 42.50%
II
HELIX: 0.83% BS: 28.06%
COIL: 24.17% TURN: 46.94%
III
HELIX: 0.83% BS: 31.39%
COIL: 25.83% TURN: 41.94%
IV
HELIX: 0% BS: 31.48%
COIL: 26.76% TURN: 41.76%
V
HELIX: 0.28% BS: 28.80%
COIL: 25.56% TURN: 45.37%

Figure 4

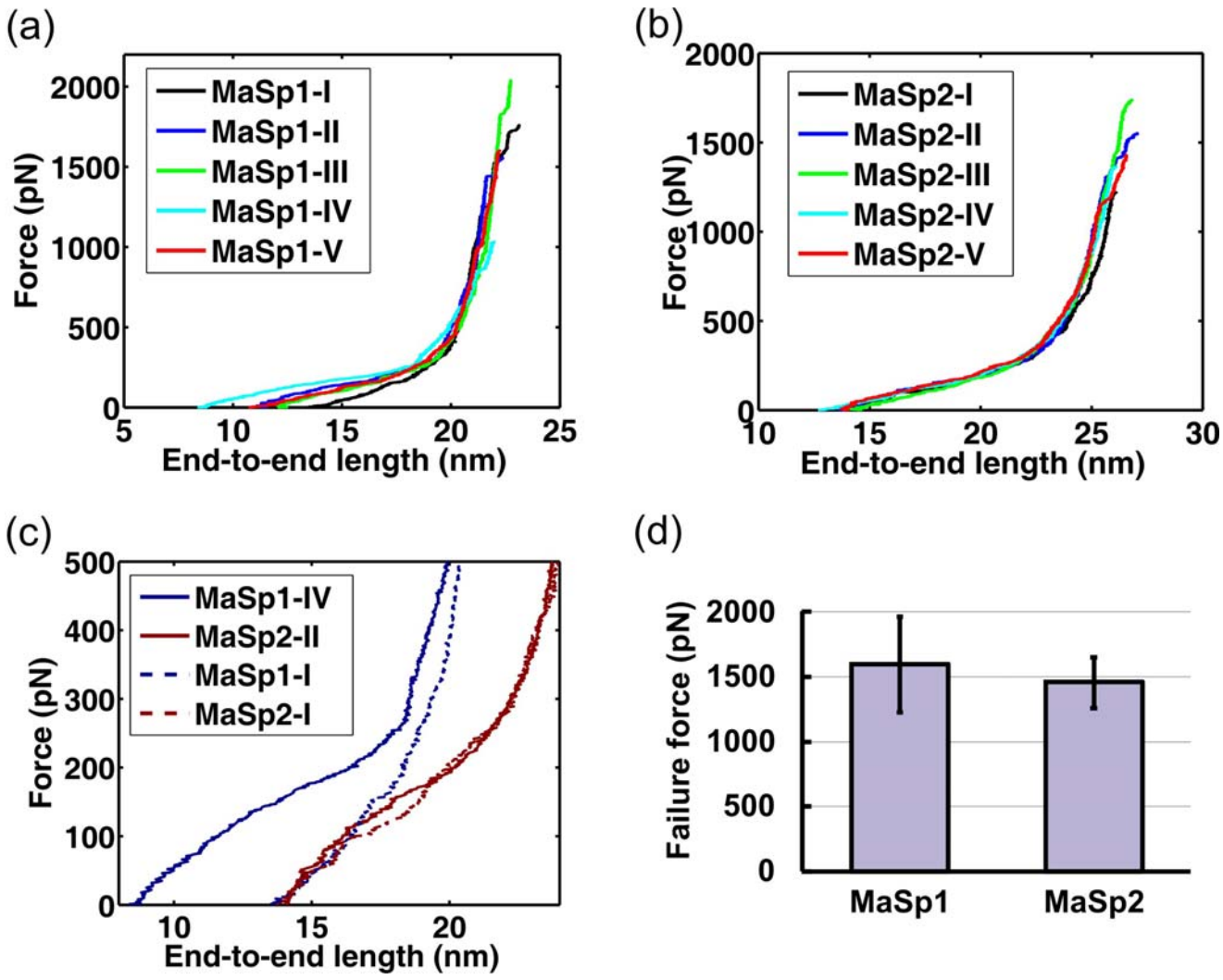
(a)
MaSp1-I
MaSp1-II
MaSp1-III
MaSp1-IV
MaSp1-V
Force (pN)
End-to-end length (nm)
(b)
MaSp2-I
MaSp2-II
MaSp2-III
MaSp2-IV
MaSp2-V
Force (pN)
End-to-end length (nm)
(c)
MaSp1-IV
MaSp2-II
MaSp1-I
MaSp2-I
Force (pN)
End-to-end length (nm)
(d)
Failure force (pN)
MaSp1
MaSp2

Figure 5

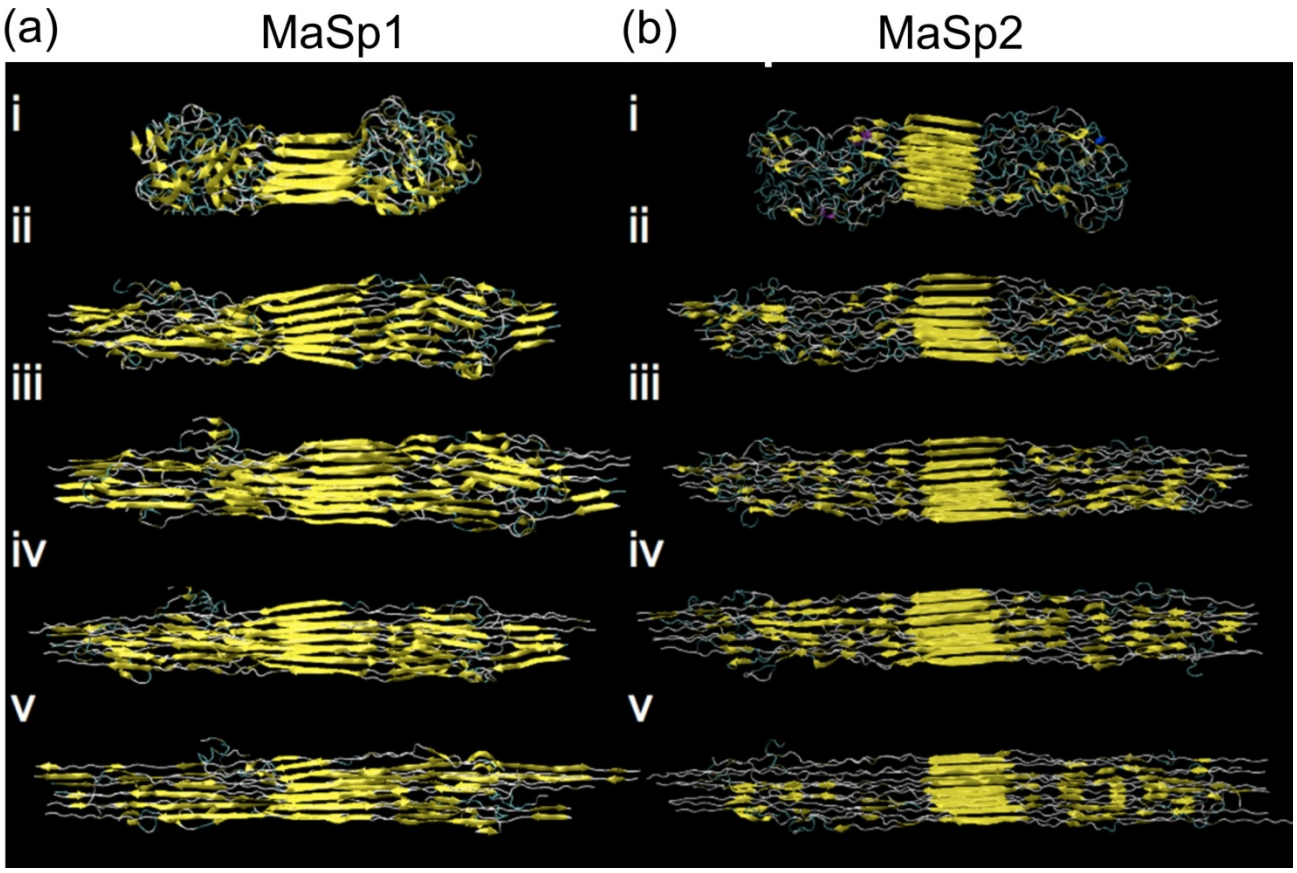

Figure 6

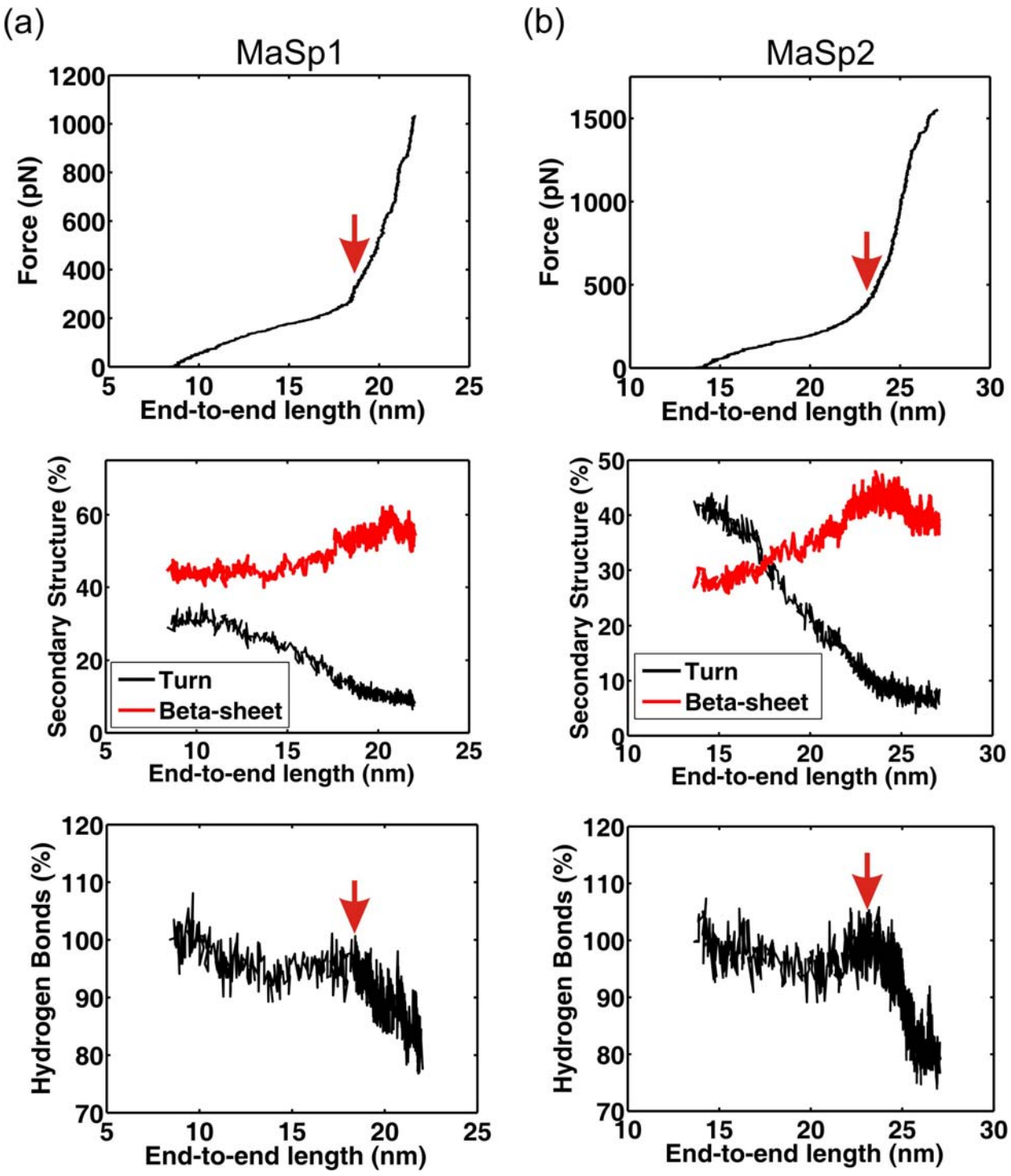
(a)
MaSp1
Force (pN)
End-to-end length (nm)
Secondary Structure (%)
Turn
Beta-sheet
Hydrogen Bonds (%)
(b)
MaSp2

Figure 7